

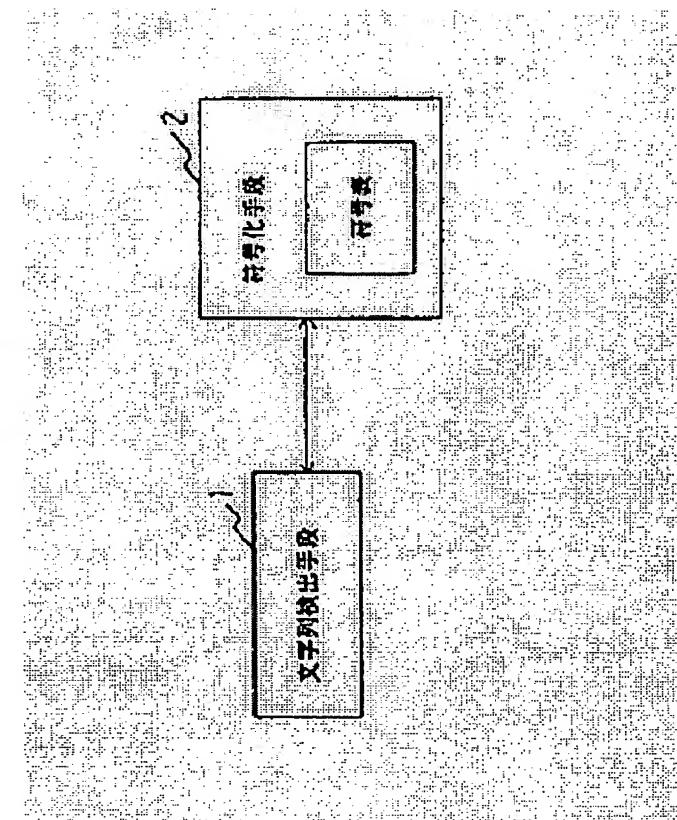
COMPRESSION SYSTEM FOR INFORMATION ON JAPANESE WORD CHARACTER STRING

Patent number: JP2255977
Publication date: 1990-10-16
Inventor: KOSAKAI TAKAO
Applicant: NEC CORP
Classification:
- International: G06F15/20
- European:
Application number: JP19890021210 19890130
Priority number(s):

Abstract of JP2255977

PURPOSE: To improve the compressing efficiency of the Japanese word information which is repetitively used by the comparatively long words of KANJI (Chinese character) or KATAKANA (square form of Japanese syllabary) by using a means which detects the continuous groups of Japanese word processes of the specific codes and a means which allocates the codes to these continuous groups.

CONSTITUTION: A character string detecting means 1 sets a process pointer at the head of the Japanese word information and calls out an encoding means 2 to carry out an encoding operation when the former half bytes of a character of a character string the equal to the sexadecimal 30 - 7E or 25. The means 2 extracts the continuous characters of a character string whose former half bytes are equal to the sexadecimal 30 - 7E and the continuous characters whose former bytes are equal to sexadecimal 25 and 21 with the latter half bytes equal to a sexadecimal 3C. When the character string is registered into a code table, the relevant register number is taken out. Otherwise a character string where the value next to the register number registered into the code table right before a due time point is extracted as a register number is registered into the code table. Then the value obtained by adding the highest rank bit to the value whose register number is expressed in a binary number is stored as a converted code.



⑨日本国特許庁(JP) ⑩特許出願公開
⑪公開特許公報(A) 平2-255977

⑤Int.Cl.⁵
G 06 F 15/20

識別記号 庁内整理番号
580 L 7165-5B

⑬公開 平成2年(1990)10月16日

審査請求 未請求 請求項の数 1 (全3頁)

④発明の名称 日本語文字列情報圧縮方式

②特願平1-21210
②出願平1(1989)1月30日

⑦発明者 小堺 隆男 東京都港区芝5丁目33番1号 日本電気株式会社内
⑦出願人 日本電気株式会社 東京都港区芝5丁目7番1号
⑦代理人 弁理士 内原 晋

明細書

1. 発明の名称

日本語文字列情報圧縮方式

2. 特許請求の範囲

2バイトで表現される日本語コードで表現された日本語情報を圧縮する処理において、特定のコードの値の日本語処理の連続した塊を検出する文字列検出手段と、その塊に符号を割り当てる符号化手段とを具備することを特徴とする日本語文字列情報圧縮方式。

3. 発明の詳細な説明

〔産業上の利用分野〕

本発明は、JIS-X-0208“情報交換用漢字符号系”に代表される通常の2バイトコードで表現された日本語情報を圧縮する日本語文字列情報圧縮方式に関する。

〔従来の技術〕

従来、日本語情報を圧縮する場合、日本語文字が14ビットで表され、2バイトに格納するときにできる2ビットの余りを詰める方法が通常とられている。

〔発明が解決しようとする課題〕

しかしながら、前述の圧縮方法では約1.14倍の圧縮効果しか得られない。

本発明の目的は、これを大きく上回る圧縮率を実現する日本語文字列情報圧縮方式を提供することにある。

〔課題を解決するための手段〕

本発明の日本語文字列情報圧縮方式は、2バイトで表現される日本語コードで表現された日本語情報を圧縮する処理において、特定のコードの値の日本語処理の連続した塊を検出する文字列検出手段と、その塊に符号を割り当てる符号化手段とを具備することを特徴とする。

〔実施例〕

次に本発明について図面を参照して説明する。

第1図は本発明の一実施例のブロック図であり、この実施例は、文字列検出手段1と、符号化手段2とから構成されている。

次に、文字列検出手段1における動作を詳細に説明する。第2図のステップ21で処理と対象となる日本語情報の先頭に処理ポインタを位置付け、ステップ22で日本語文字列の文字列の文字の前半バイトが16進数の30～7Eかまたは25であれば符号化手段2を呼び出して符号化を行う。そうでなければステップ24でそのまま格納する(圧縮しない)。ステップ25では処理を終了するか否かを判定する。

次に、第1図の符号化手段2における動作を詳細に説明する。第3図のステップ31で日本語文字列の文字の前半バイトが16進数の30～7Eである文字の連なり、または前半バイトが16進数の25および前半バイトが16進数の21後半バイトが16進数の3Cである文字の連なりを抽出する。ステップ32で抽出された文字列が符号表で既に登録されていれば、ステップ33で符号

表から登録番号を取り出す。そうでなければステップ34で直前に符号表に登録した登録番号の次の値を登録番号として、抽出された文字列を符号表に登録する。ステップ35で登録番号を2進数で表した値に最上位のビットをオンにしたもの変換符号として格納する。

〔発明の効果〕

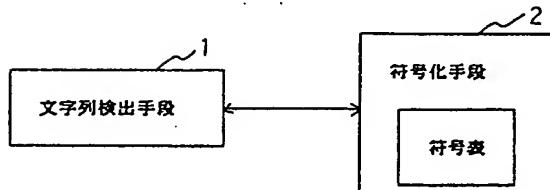
以上説明したように本発明には、比較的長い漢字列や片仮名の単語が繰り返し使用される日本語情報の圧縮において特に高い圧縮効果がある。

4. 図面の簡単な説明

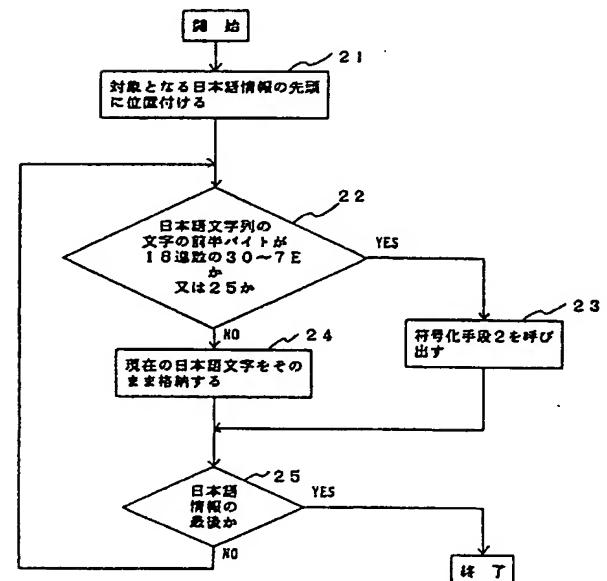
第1図は本発明の一実施例のブロック図、第2図は第1図中の文字列検出手段の動作を示す流れ図、第3図は第1図中の符号化手段の動作を示す流れ図である。

1…文字列検出手段、2…符号化手段。

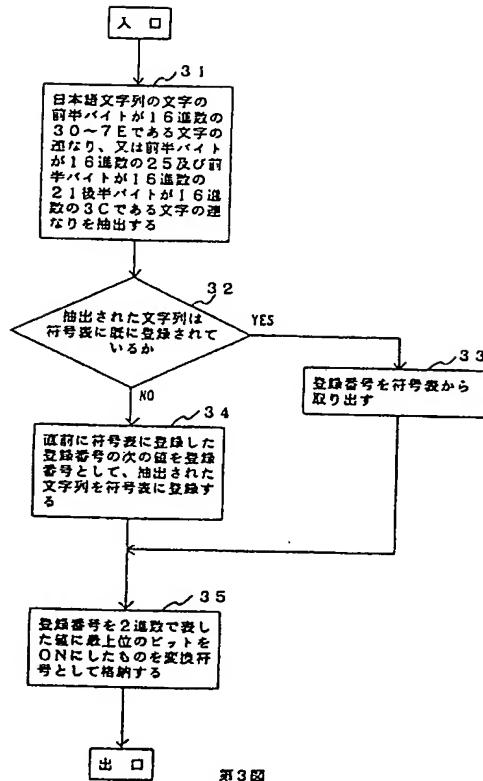
代理人 弁理士 内原 晋



第1図



第2図



第3図